WYC:dks

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Box PATENT APPLICATION
TO THE ASSISTANT COMMISSIONER FOR PATENTS
Washington, D.C. 20231

Transmitted herewith for filing is the patent application of:

Inventor(s): William Y. Conwell

For: METHODS OF PROCESSING TEXT FOUND IN IMAGES

Enclosed are:

☒   3 pages of specification, 1 pages of claims, an abstract and a Combined Declaration and Power of Attorney.

☒   1 sheet(s) of formal drawings.

☒   An assignment of the invention to: Digimarc Corporation and a Recordation Cover Sheet.

For	Claims filed	FILING FEE Number Allotted		Number Extra	Rate	Basic Fee $690.00
Total Claims	4	20	=	0	$18.00	0
Independent Claims	2	3	=	0	$78.00	0
TOTAL FILING FEE			=			$690.00

☒   Please charge the filing fee of **$690.00** and the assignment recordal fee of **$40.00** and any additional fees which may be required in connection with the filing of this application and recording any assignment filed herewith, or credit over-payment, to Account No. 50-1071. A copy of this sheet is enclosed.

☒   Please return the enclosed postcard to confirm that the items listed above have been received.

Respectfully submitted,

DIGIMARC CORPORATION

Date: September 26, 2000

Digimarc Corporation
19801 SW 72nd Avenue, Suite 250
Tualatin, OR 97062
Phone: 503-885-8699

By
William Y. Conwell
Registration No. 31,943

METHODS OF PROCESSING TEXT FOUND IN IMAGES

The World Wide Web is a distributed database including hundreds of millions of documents. Search engines such as Alta Vista attempt to index the web based on ASCII

5   text included on each page and on associated meta tags. Increasingly, however, text information is present on the Web in the form of text images. Known search engines are unable to make use of text presented in this form.

One approach to this problem is discussed in Lopresti et al, "Locating and Recognizing Text in WWW Images," Information Retrieval, vol.2, no.2-3 p.177-206,

10   2000, and involves a procedure based on clustering in color space followed by a connected-components analysis. Character recognition is performed using polynomial surface fitting and "fuzzy" n-tuple classifiers. While suitable for some applications, such techniques are too computationally intensive and imprecise for widespread use.

In accordance with one embodiment of the present invention, an image containing

15   text is digitally watermarked with an identifier. The identifier serves as an index to a database record where additional information about the image, including keywords or full text of the included text, are provided. To obtain the associated data, a search engine web crawler or other process can download an image, apply a watermarking detection procedure, use an identifier thereby obtained to index a database, and access keywords or

20   full text represented in the image from the indexed database record.

The text can be entered in the database using various known methods. One is to have the text manually coded by a clerical service. Another is to apply an automated OCR process to the image data, such as that detailed by Lopresti. Once the text is once thereby developed, it can be made quickly available repeatedly thereafter by reference to

25   the associated database record.

The database can be conventional, and is preferably accessible over the internet. A suitable database system is disclosed in copending application 09/571,422, filed May 15, 2000. A variety of watermarking techniques are known. An illustrative set of techniques that can be employed in this application is disclosed in copending application

09/503,881, filed February 14, 2000. The disclosures of these applications are incorporated herein by reference.

The technology disclosed herein finds myriad applications. As noted, one is in the indexing of a collection of electronic documents (e.g., web pages). An index augmented by the results of such a procedure is generally more useful than such an index without augmentation.

Another application is in the use of webcams, or security monitoring cameras. Certain image frames from such sources (e.g., one every minute, or one every second, etc.) can be analyzed for textual information (e.g., license plate markings, superimposed date data), and the textual information stored. The image data is watermarked, with the watermark indicating the repository of the corresponding textual information.

Still another application is PDF documents or fax data files. (While some PDF files include corresponding ASCII text data, most do not.) The file data can be applied to an OCR engine, and the resulting text stored in a database. The PDF or fax data file can be slightly altered to impart a watermark – the watermark again serving to point to the repository of the corresponding text information.

Yet another application is in photocopiers. Again, the textual content is extracted from the scanned image of the original document. In this case the paper photocopy output (or a corresponding digital file) is altered in slight respects to encode a watermark. The watermark points to the text data repository.

While the illustrative embodiment particularly considered watermarks that convey an index to a remote database, other arrangements are naturally possible. For example, the watermark can directly encode the fulltext or keywords (forms of metadata).

Similarly, while the illustrative embodiment particularly considered imaged text in image files, the same principles can be applied more widely. For example, all metadata associated with an image through a watermark can be employed in compiling an index of the web or other collection of content data – not just included text (e.g., names of persons and places, dates, times, and other more application-specific metadata). Moreover, such techniques are not just limited to images. Other forms of content,

including video and audio, can be watermarked, and the metadata thereby associated with the content can be used for web indexing and other purposes.

I CLAIM

1. A method comprising:

receiving data corresponding to an image, the image including a depiction of text;

decoding a digital watermark from the image data; and

5   by reference to said digital watermark, accessing at least some of said depicted

text in non-image form.

2. An index to a collection of electronic objects, at least one of said objects comprising an image depicting text, formed by use of the method of claim 1.

10

3. A method comprising:

receiving data corresponding to an image, the image including a depiction of text;

generating a non-image representation of at least some of said depicted text;

encoding a watermark in a representation of said image; and

15   associating with said watermark

4. The method of claim 2 in which said non-image representation comprises ASCII text.

COMBINED DECLARATION AND POWER OF ATTORNEY
FOR PATENT APPLICATION

As a below named inventor, I hereby declare that:

My residence, post office address and citizenship are as stated below next to my name,
I believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and
joint inventor (if plural names are listed below) of the subject matter which is claimed and for which a
patent is sought on the invention entitled METHODS OF PROCESSING TEXT FOUND IN IMAGES, the
specification of which

[x]   is attached hereto.

I hereby state that I have reviewed and understand the contents of the above-identified
specification, including the claims, as amended by any amendment referred to above.

I acknowledge the duty to disclose information which is material to patentability as defined in Title
37, Code of Federal Regulations, § 1.56. If this is a continuation-in-part application filed under the
conditions specified in 35 U.S.C. § 120 which discloses and claims subject matter in addition to that
disclosed in the prior copending application, I further acknowledge the duty to disclose material
information as defined in 37 CFR § 1.56 which occurred between the filing date of the prior application and
the national or PCT international filing date of the continuation-in-part application.

I hereby claim foreign priority benefits under Title 35, United States Code, § 119(a)-(d) of any
foreign application(s) for patent or inventor's certificate or of any PCT International application(s)
designating at least one country other than the United States of America listed below and have
also identified below any foreign application(s) for patent or inventor's certificate or any PCT International
application(s) designating at least one country other than the United States of America filed by me on the
same subject matter having a filing date before that of the application(s) on which priority is claimed:

Prior Foreign Application(s)

Priority
Claimed

(Number)	(Country)	(Day/Month/Year Filed)	[ ] Yes	[ ] No

I hereby claim the benefit under Title 35, United States Code, § 119(e) of any United States
provisional application(s) listed below:

Application Number	Filing Date

I hereby claim the benefit under Title 35, United States Code, § 120 of any United States
application(s) or § 365(c) of any PCT International application(s) designating the United States, listed
below and, insofar as the subject matter of each of the claims of this application is not disclosed in the
prior United States or PCT International application in the manner provided by the first paragraph of Title
35, United States Code, § 112, I acknowledge the duty to disclose material information as defined in Title
37, Code of Federal Regulations, § 1.56(a) which occurred between the filing date of the prior application
and the national or PCT International filing date of this application:

(Application No.)	(Filing Date)	(Status: patented, Pending, abandoned)

The undersigned hereby authorizes the U.S. attorney or agent named herein to accept and follow instructions from _____ as to any action to be taken in the Patent and Trademark Office regarding this application without direct communication between the U.S. attorney or agent and the undersigned. In the event of a change in the persons from whom instructions may be taken, the U.S. attorney or agent named herein will be so notified by the undersigned.

I hereby appoint the following attorney(s) and/or agent(s) to prosecute this application, to file a corresponding international application, and to transact all business in the Patent and Trademark Office connected therewith:

William Y. Conwell	Reg. No. 31,943
Joel R. Meyer	Reg. No. 37,677
Thomas M. Horgan	Reg. No. 33,183
Elmer Galbi	Reg. No. 19,761

Address all telephone calls to William Y. Conwell at telephone number (503) 968-0443. Address all correspondence to:
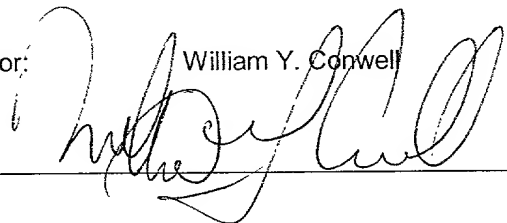
William Y. Conwell
DIGIMARC Corporation
Digimarc Corporation
19801 SW 72nd Avenue, Suite 250
Tualatin, OR 97062

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

Full Name of First or Sole Inventor: William Y. Conwell

Inventor's Signature _____   _____
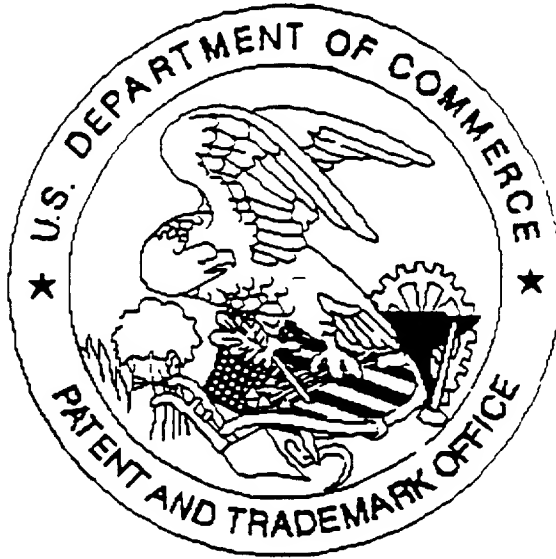   Date

Residence: Portland, Oregon

Citizenship: USA

Post Office Address: 6224 SW Tower Way, Portland, OR 97221

United States Patent & Trademark Office
Office of Initial Patent Examination -- Scanning Division

SCANNED, #_____

Application deficiencies were found during scanning:

☑ Page(s) ___1___ of ___drawing___ were not present
for scanning.
(Document title)

☐ Page(s) _____ of _____ were not present
for scanning.
(Document title)

☐ Scanned copy is best available.